University of Bialystok

YARS-PG: Serialization for Property Graphs

Dominik Tomaszuk[1]   Renzo Angles[2,3]   Łukasz Szeremeta[1]   Karol Litman[1]   Diego Cisterna[3]

[1]Institute of Informatics,
University of Bialystok, Bialystok, Poland

[2]Department of Computer Science,
Universidad de Talca, Curicó, Chile

[3]Center for Semantic Web Research,
Santiago, Chile

Abstract

Graph serialization is critical for the evolution of graph-oriented systems and applications. Unfortunately, there is still no universal serialization for property graphs that will cover all main features of graph databases systems. We propose YARS-PG property graphs serialization which is simple, extensible, and platform independent. YARS-PG supports all the features allowed by the current database systems based on the property graph data model, and can be adapted in the future to work with various database systems, visualization software and other graph-oriented tools.

Introduction

Data serialization is fundamental in data management as it allows to support database exchange, systems benchmarking, data visualization and presentation. Data serialisation is very important for several reasons: it simplifies translation into other data formats; it enables automatic data processing; makes easier to compare databases, because the same data can be shared between systems; it is essential to foster the interoperability of heterogeneous databases; it results in a simpler backup method; other processing and visualization tools can read the data.

Position Statement

Graph serialization is nothing new. The lack of a specific Property Graph [4] data model directly affects the development of other components, including query languages and serialization formats. Although there are some graph data syntaxes available (like GraphML [1] or GEXF [2]), there is no a standard one, and none of them is able to cover all the features presented by the property graph data model.

System	Properties			Labels		Edges					Format			
	Pairs	Multiple	*Null* value	Multiple	Unique	Directed	Undirected	Multiple	Same label	Unstructured	XML	JSON	Textual	Tabular
GEXF	•									•	•			
GDF			∘△			•		•	∘▲					•
GML	•		∘△			•	∘▽	•					•	
GraphML	•	∘			•▼	•		•			•		•	
Pajek NET				•		•	•	•						•
GraphViz DOT	•	∘	∘△	•		•	•	•	•				•	
UCINET DL	•					•	•	•					•	
Tulip TPL		•				•							•	
Netdraw VNA	•		∘△		•	•		•						•
DotML				•		•		•	•		•			
S-Dot				•		•		•	•				•	
GraphSON TP2						•		•	•			•		
GraphSON TP3	•	∘		•	•	•		•	•			•		
DGML						•		•	•		•		•	
GXL	•	∘			•▼	•		•			•			
YARS-PG	•	•	•	•	∘	•	•	•	•	•			•	

△ no grammar
▽ only global definition
▲ labels supported as properties
▼ only in the sense of identifiers

Comparison of Property Graph serializations

Property Graphs should have a well-defined serialization in order to convert data (obtained from a source system) into a syntax that can be stored (in the same system) or forwarded (to a target system), and then constructed again.

Contact Information

- Dominik Tomaszuk – d.tomaszuk@uwb.edu.pl
- Renzo Angles – rangles@utalca.cl
- Łukasz Szeremeta – l.szeremeta@uwb.edu.pl
- Karol Litman – karolks94@gmail.com
- Diego Cisterna – dcisterna@live.com

Our proposal

Understanding the use cases and requirements of different people will help define the needs and the priority of features. Finally, on fundamentals of this, we propose YARS-PG [5] that is platform independent and extensible serialization. Our serialization supports all the features allowed by the current database systems based on the property graph data model, and can be adapted to work with various visualization software, database-driven systems and other graph-oriented tools.

```
1  <"Author01">{"Author"}["fname":"John","lname":"Smith"]
2  <"Author02">{"Author"}["fname":"Alice","lname":"Brown"]
3  <"EI01">{"Entry","InProc"}["title":"Serialization for...",
      "numpages":10,"keyword":"Graph database"]
4  <"EA01">{"Entry","Article"}["title":"Property Graph...",
      "numpages":10,"keyword":["Query", "Graph"]]
5  <"Proc01">{"Proceedings"}["title":"BDAS","year":2018,"month":"May"]
6  <"Jour01">{"Journal"}["title":"J. DB","year":2020,"vol":30]
7
8  ("EI01")-{"has_author"}["order":1]->("Author01")
9  ("EI01")-{"has_author"}["order":2]->("Author02")
10 ("EA01")-{"has_author"}["order":1]->("Author02")
11 ("EA01")-{"cites"}->("EI01")
12 ("EI01")-{"booktitle"}["pages":"111-121"]->("Proc01")
13 ("EA01")-{"published_in"}["pages":"222-232"]->("Jour01")
```

YARS-PG example

The entire YARS-PG grammar in ANTLR 4 [3] and also in EBNF notation[a] has been made available in the GitHub repository [5].

Conclusions

Serialization methods are fundamental in graph data management to support database exchange, benchmarking of systems, and data visualization. We present YARS-PG, a data format for serializing property graphs. YARS-PG was designed to be simple, extensible and platform independent, and to support all the features provided by the current database systems based on the property graph data model.
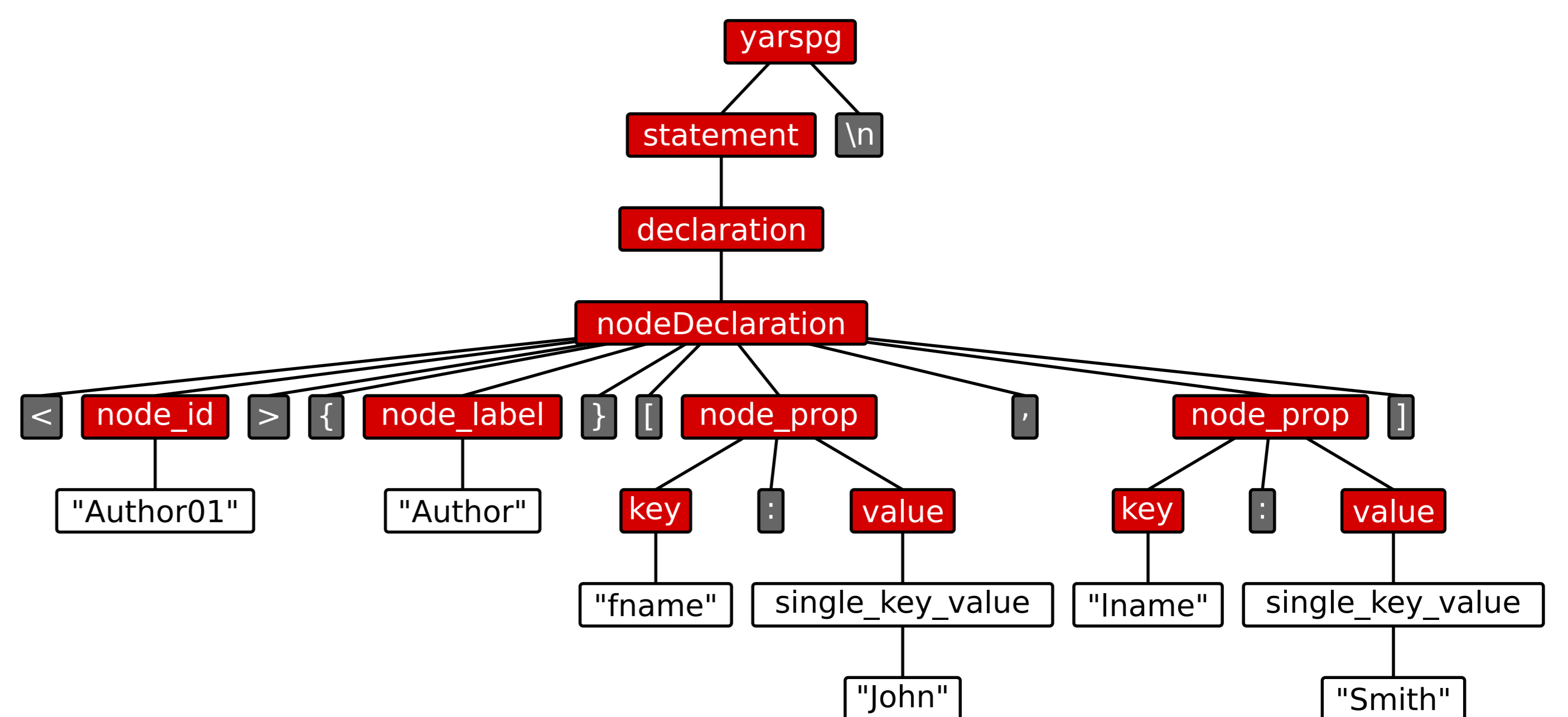
Parse tree fragment of first line in example

References

[1] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M. S. Marshall. GraphML Progress Report Structural Layer Proposal. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Graph Drawing*, pages 501–512, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-45848-7. doi: 10.1007/3-540-45848-4_59.

[2] S. Heymann. *Gephi*, pages 612–625. Springer New York, New York, NY, 2014. ISBN 978-1-4614-6170-8. doi: 10.1007/978-1-4614-6170-8_299.

[3] T. Parr. *The definitive ANTLR 4 reference*. Pragmatic Bookshelf, 2013.

[4] D. Tomaszuk and Ł. Szeremeta. Named property graphs. In *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 173–177. IEEE, 2018. doi: 10.15439/2018F103.

[5] Łukasz Szeremeta. lszeremeta/antlr-yarspg: Yeah - YARS-PG 3.0.0!, Mar. 2019. URL https://doi.org/10.5281/zenodo.2581261.

[a]https://www.w3.org/TR/REC-xml/#sec-notation